Description

CHIMERIC GENE CONSTRUCTS FOR GENERATION OF FLUORESCENT TRANSGENIC ORNAMENTAL FISH

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This is a continuation of co-pending application Serial No. 09/913,898, filed October 3, 2001, which is a national-ization of PCT application WO 00/49150 filed July 16, 1999, claiming priority over a Singapore application filed July 14, 1999, and an earlier Singapore application, Serial No. 9900811-2, filed February 18, 1999, all of which are incorporated herein by reference in their entirety.

BACKGROUND OF INVENTION

[0002] This invention relates to fish gene promoters and chimeric gene constructs with these promoters for generation of transgenic fish, particularly fluorescent transgenic orna-mental fish.

[0003] Transgenic technology involves the transfer of a foreign

gene into a host organism enabling the host to acquire a new and inheritable trait. The technique was first developed in mice by Gordon et al. (1980). They injected foreign DNA into fertilized eggs and found that some of the mice developed from the injected eggs retained the foreign DNA. Applying the same technique, Palmiter et al. (1982) have introduced a chimeric gene containing a rat growth hormone gene under a mouse heavy metal-inducible gene promoter and generated the first batch of genetically engineered supermice, which are almost twice as large as non-transgenic siblings. This work has opened a promising avenue in using the transgenic approach to provide to animals new and beneficial traits for livestock husbandry and aquaculture.

[0004] In addition to the stimulation of somatic growth for increasing the gross production of animal husbandry and aquaculture, transgenic technology also has many other potential applications. First of all, transgenic animals can be used as bioreactors to produce commercially useful compounds by expression of a useful foreign gene in milk or in blood. Many pharmaceutically useful protein factors have been expressed in this way. For example, human αl-antitrypsin, which is commonly used to treat emphysema,

has been expressed at a concentration as high as 35 mg/ml (10% of milk proteins) in the milk of transgenic sheep (Wright et al., 1991). Similarly, the transgenic technique can also be used to improve the nutritional value of milk by selectively increasing the levels of certain valuable proteins such as caseins and by supplementing certain new and useful proteins such as lysozyme for antimicrobial activity (Maga and Murray, 1995). Second, transgenic mice have been widely used in medical research, particularly in the generation of transgenic animal models for human disease studies (Lathe and Mullins, 1993). More recently, it has been proposed to use transgenic pigs as organ donors for xenotransplantation by expressing human regulators of complement activation to prevent hyperacute rejection during organ transplantation (Cozzi and White, 1995). The development of disease resistant animals has also been tested in transgenic mice (e.g. Chen et al., 1988).

[0005] Fish are also an intensive research subject of transgenic studies. There are many ways of introducing a foreign gene into fish, including: microinjection (e.g. Zhu et al., 1985; Du et al., 1992), electroporation (Powers et al., 1992), sperm-mediated gene transfer (Khoo et al., 1992;

Sin et al., 1993), gene bombardment or gene gun (Zelemin et al., 1991), liposome-mediated gene transfer (Szelei et al., 1994), and the direct injection of DNA into muscle tissue (Xu et al., 1999). The first transgenic fish report was published by Zhu et al. (1985) using a chimeric gene construct consisting of a mouse metallothionein gene promoter and a human growth hormone gene. Most of the early transgenic fish studies have concentrated on growth hormone gene transfer with an aim of generating fast growing "superfish". A majority of early attempts used heterologous growth hormone genes and promoters and failed to produce gigantic superfish (e.g. Chourrout et al., 1986; Penman et al., 1990; Brem et al., 1988; Gross et al., 1992). But enhanced growth of transgenic fish has been demonstrated in several fish species including Atlantic salmon, several species of Pacific salmons, and loach (e.g. Du et al., 1992; Delvin et al., 1994, 1995; Tsai et al., 1995).

[0006] The zebrafish, Danio rerio, is a new model organism for vertebrate developmental biology. As an experimental model, the zebrafish offers several major advantages such as easy availability of eggs and embryos, tissue clarity throughout embryogenesis, external development, short

generation time and easy maintenance of both the adult and the young. Transgenic zebrafish have been used as an experimental tool in zebrafish developmental biology. However, despite the fact that the first transgenic zebrafish was reported a decade ago (Stuart et al., 1988), most transgenic zebrafish work conducted so far used heterologous gene promoters or viral gene promoters: e.g. viral promoters from SV40 (simian virus 40) and RSV (Rous sarcoma virus) (Stuart et al., 1988, 1990; Bayer and Campos-Ortega, 1992), a carp actin promoter (Liu et al., 1990), and mouse homeobox gene promoters (Westerfield et al., 1992). As a result, the expression pattern of a transgene in many cases is variable and unpredictable.

[0007] GFP (green fluorescent protein) was isolated from a jelly fish, Aqueous victoria. The wild type GFP emits green fluorescence at a wavelength of 508 nm upon stimulation with ultraviolet light (395 nm). The primary structure of GFP has been elucidated by cloning of its cDNA and genomic DNA (Prasher et al., 1992). A modified GFP, also called EGFP (Enhanced Green Fluorescent Protein) has been generated artificially and it contains mutations that allow the protein to emit a stronger green light and its coding sequence has also been optimized for higher ex-

pression in mammalian cells based on preferable human codons. As a result, EGFP fluorescence is about 40 times stronger than the wild type GFP in mammalian cells (Yang et al., 1996). GFP (including EGFP) has become a popular tool in cell biology and transgenic research. By fusing GFP with a tested protein, the GFP fusion protein can be used as an indicator of the subcellular location of the tested protein (Wang and Hazelrigg, 1994) . By transformation of cells with a functional GFP gene, the GFP can be used as a marker to identify expressing cells (Chalfie et al., 1994). Thus, the GFP gene has become an increasingly popular reporter gene for transgenic research as GFP can be easily detected by a non-invasive approach.

[0008] The GFP gene (including EGFP gene) has also been intro-duced into zebrafish in several previous reports by using various gene promoters, including Xenopus elongation factor 1α enhancer-promoter (Amsterdam et al., 1995, 1996), rat myosin light-chain enhancer (Moss et al., 1996), zebrafish GATA-1 and GATA-3 promoters (Meng et al., 1997; Long et al., 1997), zebrafish α- and β-actin promoters (Higashijima et al., 1997), and tilapia insulin-like growth factor I promoter (Chen et al., 1998). All of these transgenic experiments aim at either developing a

GFP transgenic system for gene expression analysis or at testing regulatory DNA elements in gene promoters.

SUMMARY OF INVENTION

[0009] It is a primary objective of the invention to clone fish gene promoters that are constitutive (ubiquitous), or that have tissue specificity such as skin specificity or muscle specificity or that are inducible by a chemical substance, and to use these promoters to develop effective gene constructs for production of transgenic fish.

[0010] It is another objective of the invention to develop fluorescent transgenic ornamental fish using these gene constructs. By applying different gene promoters, tissue-specific, inducible under different environmental conditions, or ubiquitous, to drive the GFP gene, GFP could be expressed in different tissues or ubiquitously. Thus, these transgenic fish may be skin fluorescent, muscle fluorescent, ubiquitously fluorescent, or inducibly fluorescent. These transgenic fish may be used for ornamental purposes, for monitoring environmental pollution, and for basic studies such as recapitulation of gene expression programs or monitoring cell lineage and cell migration. These transgenic fish may be used for cell transplantation and nuclear transplantation or fish cloning.

[0011]   Other objectives, features and advantages of the present invention will become apparent from the detailed description which follows, or may be learned by practice of the invention.

[0012]   Four zebrafish gene promoters of different characteristics were isolated and four chimeric gene constructs containing a zebrafish gene promoter and EGFP DNA were made: pCK-EGFP, pMCK-EGFP, pMLC2f-EGFP and pARP-EGFP. The first chimeric gene construct, pCK-EGFP, contains a 2.2 kbp polynucleotide comprising a zebrafish cytokeratin (CK) gene promoter which is specifically or predominantly expressed in skin epithelia. The second one, pMCK-EGFP, contains a 1.5 kbp polynucleotide comprising a muscle-specific promoter from a zebrafish muscle creatine kinase (MCK) gene and the gene is only expressed in the muscle tissue. The third construct, pMLC2f-EGFP contains a 2.2 kpb polynucleotide comprising a strong skeletal muscle-specific promoter from the fast skeletal muscle isoform of the myosin light chain 2 (MLC2f) gene and is expressed specifically or predominantly in skeletal muscle. The fourth chimeric gene construct, pARP-EGFP, contains a strong and ubiquitously expressed promoter from a zebrafish acidic ribosomal protein (ARP) gene. These four

chimeric gene constructs have been introduced into zebrafish at the one cell stage by microinjection. In all cases, the GFP expression patterns were consistent with the specificities of the promoters. GFP was predominantly expressed in skin epithelia with pCK-EGFP, specifically expressed in muscles with pMCK-EGFP, specifically expressed in skeletal muscles with pMLC2f-EGFP and ubiquitously expressed in all tissues with pARP-EGFP.

[0013] These chimeric gene constructs are useful to generate green fluorescent transgenic fish. The GFP transgenic fish emit green fluorescence light under a blue or ultraviolet light and this feature makes the genetically engineered fish unique and attractive in the ornamental fish market. The fluorescent transgenic fish are also useful for the development of a biosensor system and as research models for embryonic studies such as cell lineage, cell migration, cell and nuclear transplantation etc.

BRIEF DESCRIPTION OF DRAWINGS

[0014] Figs. 1A-1I are photographs showing expression of CK (Figs. 1A-1C), MCK (Figs. 1D-1E), ARP (Figs. 1F-1G) and MLC2f (Figs. 1H-1I) mRNAs in zebrafish embryos as revealed by whole mount in situ hybridization (detailed description of the procedure can be found in Thisse et al.,

1993). (Fig. 1A) A 28 hpf (hour postfertilization) embryo hybridized with a CK antisense riboprobe. (Fig. 1B) Enlargement of the mid-part of the embryo shown in Fig. 1A. (Fig. 1C) Cross-section of the embryo in Fig. 1A. (Fig. 1D) A 30 hpf embryo hybridized with an MCK antisense riboprobe. (Fig. 1E) Cross-section of the embryo in Fig 1D. (Fig. 1F) A 28 hpf embryo hybridized with an ARP antisense riboprobe. (Fig. 1G) Cross-section of the embryo in Fig. 1F. Arrows indicate the planes for cross-sections and the box in panel A indicates the enlarged region shown in panel B. (Fig. 1H) Side view of a 22-hpf embryo hybridized with the MLC2f probe. (Fig. 1I) Transverse section through the trunk of a stained 24-hpf embryo. SC, spinal cord; N, notochord.

[0015] Fig. 2A is a digitized image showing distribution of CK, MCK and ARP mRNAs in adult tissues. Total RNAs were prepared from selected adult tissues as indicated at the top of each lane and analyzed by Northern blot hybridization (detailed description of the procedure can be found in Gong et al., 1992). Three identical blots were made from the same set of RNAs and hybridized with the CK, MCK and ARP probes, respectively.

[0016] Fig. 2B is a digitized image showing distribution of MLC2f

mRNA in adult tissues. Total RNAs were prepared from selected adult tissues as indicated at the top of each lane and analyzed by Northern blot hybridization (detailed description of the procedure can be found in Gong et al., 1992). Two identical blots were made from the same set of RNAs and hybridized with the MLC2f probe and a ubiquitously expressed ß-actin probe, respectively.

[0017] Fig. 3. is a schematic representation of the strategy of promoter cloning. Restriction enzyme digested genomic DNA was ligated with a short linker DNA which consists of Oligo 1 and Oligo 2. Nested PCR reactions were then performed: the first round PCR used linker specific primer L1 and gene specific primers G1, where G1 is CK1, MCK1, M1 or ARP1 in the described embodiments, and the second round linker specific primer L2 and gene specific primer G2, where G2 is CK2, MCK2, M2 or ARP2, respectively in the described embodiments.

[0018] Fig. 4 is a schematic map of the chimeric gene construct, pCK-EGFP. The 2.2 kb zebrafish DNA fragment comprising the CK promoter region is inserted into pEGFP-1 (Clonetech) at the EcoRI and BamHI site as indicated. In the resulting chimeric DNA construct, the EGFP gene is under control of the zebrafish CK promoter. Also shown is

the kanamycin/neomycin resistance gene (Kan^r/Neo^r) in the backbone of the original pEGFP-1 plasmid. The total length of the recombinant plasmid pCK-EGFP is 6.4 kb.

[0019] Fig. 5 is a schematic map of the chimeric gene construct, pMCK-EGFP. The 1.5 kb zebrafish DNA fragment comprising the MCK promoter region is inserted into pEGFP-1 (Clonetech) at the EcoRI and BamHI site as indicated. In the resulting chimeric DNA construct, the EGFP gene is under control of the zebrafish MCK promoter. Also shown is the kanamycin/neomycin resistance gene (Kan^r/Neo^r) in the backbone of the original pEGFP-1 plasmid. The total length of the recombinant plasmid pMCK-EGFP is 5.7 kb.

[0020] Fig. 6 is a schematic map of the chimeric gene construct, pARP-EGFP. The 2.2 kb zebrafish DNA fragment comprising the ARP promoter/1st intron region is inserted into pEGFP-1 (Clonetech) at the EcoRI and BamHI site as indicated. In the resulting chimeric DNA construct, the EGFP gene is under control of the zebrafish ARP promoter. Also shown is the kanamycin/neomycin resistance gene (Kan^r/Neo^r) in the backbone of the original pEGFP-1 plasmid. The total length of the recombinant plasmid pARP-EGFP is 6.4 kb.

[0021] Fig. 7 is a schematic map of the chimeric gene construct,

pMLC2f-EGFP. The 2.0 kb zebrafish DNA fragment comprising the MLC2f promoter region is inserted into pEGFP-1 (Clonetech) at the HindIII and BamHI site as indicated. In the resulting chimeric DNA construct, the EGFP gene is under control of the zebrafish MLC2f promoter. Also shown is the kanamycin/neomycin resistance gene (Kan[r]/Neo[r]) in the backbone of the original pEGFP-1 plasmid. The total length of the recombinant plasmid pMLC2f-EGFP is 6.2 kb.

[0022] Fig. 8 is a photograph of a typical transgenic zebrafish fry (4 days old) with pCK-EGFP, which emits green fluorescence from skin epithelia under a blue light.

[0023] Fig. 9 is a photograph of a typical transgenic zebrafish fry (3 days old) with pMCK-EGFP, which emits green fluorescence from skeletal muscles under a blue light.

[0024] Fig. 10 is a photograph of a typical transgenic zebrafish fry (2 days old) with pARP-EGFP, which emits green fluorescence under a blue light from a variety of cell types such as skin epithelia, muscle cells, lens, neural tissues, notochord, circulating blood cells and yolk cells.

[0025] Figs. 11A-11B. Photographs of a typical transgenic zebrafish founder with pMLC2f-EGFP (Fig. 11A) and an F1 stable transgenic offspring (Fig. 11B). Both pictures were

taken under an ultraviolet light (365 nm). The green fluorescence can be better observed under a blue light with an optimal wavelength of 488 nm.

[0026] Figs. 12A–12C. Examples of high, moderate and low expression of GFP in transiently transgenic embryos at 72 hpf. (Fig. 12A) High expression, GFP expression was detected in essentially 100% of the muscle fibers in the trunk. (Fig. 12B) Moderate expression, GFP expression was detected in several bundles of muscle fibers, usually in the mid-trunk region. (Fig. 12C) Low expression, GFP expression occurred in dispersed muscle fibers and the number of GFP positive fibers is usually less than 20 per embryo.

[0027] Fig. 13. Deletion analysis of the MLC2f promoter in transient transgenic zebrafish embryos. A series of 5" deletions of MLC2f-EGFP constructs containing 2011–bp (2–kb), 1338–bp, 873–bp, 283–bp, 77–bp and 3–bp of the MLC2f promoter were generated by unidirectional deletion using the double–stranded Nested Deletion Kit from Pharmacia based on the manufacturer"s instructional manual. Each construct was injected into approximately 100 embryos and GFP expression was monitored in the first 72 hours of embryonic development. The level of GFP expression was classified based on the examples shown in

Figs. 12A–12C. Potential E-boxes and MEF2 binding sites, which are important for muscle-specific transcription (Schwarz et al., 1993; Olson et al., 1995), are indicated on the 2011-bp construct.

DETAILED DESCRIPTION

[0028] *Gene Constructs.* To develop successful transgenic fish with a predictable pattern of transgene expression, the first step is to make a gene construct suitable for transgenic studies. The gene construct generally comprises three portions: a gene promoter, a structural gene and transcriptional termination signals. The gene promoter would determine where, when and under what conditions the structural gene is turned on. The structural gene contains protein coding portions that determine the protein to be synthesized and thus the biological function. The structural gene might also contain intron sequences which can affect mRNA stability or which might contain transcription regulatory elements. The transcription termination signals consist of two parts: a polyadenylation signal and a transcriptional termination signal after the polyadenylation signal. Both are important to terminate the transcription of the gene. Among the three portions, selection of a promoter is very important for successful transgenic study,

and it is preferable to use a homologous promoter (homologous to the host fish) to ensure accurate gene activation in the transgenic host.

[0029] A promoter drives expression "predominantly"in a tissue if expression is at least 2–fold, preferably at least 5–fold higher in that tissue compared to a reference tissue. A promoter drives expression "specifically" in a tissue if the level of expression is at least 5–fold, preferably at least 10–fold higher, more preferably at least 50–fold higher in that tissue than in any other tissue.

[0030] *Recombinant DNA Constructs.* Recombinant DNA constructs comprising one or more of the DNA or RNA sequences described herein and an additional DNA and/or RNA sequence are also included within the scope of this invention. These recombinant DNA constructs usually have sequences which do not occur in nature or exist in a form that does not occur in nature or exist in association with other materials that do not occur in nature. The DNA and/or RNA sequences described as constructs or in vectors above are "operably linked" with other DNA and/or RNA sequences. DNA regions are operably linked when they are functionally related to each other. For example, DNA for a presequence or secretory leader is operably linked to

DNA for a polypeptide if it is expressed as part of a pre-protein which participates in the secretion of the polypeptide; a promoter is operably linked to a coding sequence if it controls the transcription of the coding sequence; a ribosome binding site is operably linked to a coding sequence if it is positioned so as to permit translation. Generally, operably linked means contiguous (or in close proximity to) and, in the case of secretory leaders, contiguous and in reading phase.

[0031] The sequences of some of the DNAs, and the corresponding proteins encoded by the DNA, which are useful in the invention are set forth in the attached Sequence Listing.

[0032] The complete cytokeratin (CK) cDNA sequence is shown in SEQ ID NO:1, and its deduced amino acid sequence is shown in SEQ ID NO:2. The binding sites of the gene specific primers for promoter amplification, CK1 and CK2, are indicated. The extra nucleotides introduced into CK2 for generation of a restriction site are shown as a misc_feature in the primer sequence SEQ ID NO:11. A potential polyadenylation signal, AATAAA, is indicated in SEQ ID NO:1.

[0033] The complete muscle creatine kinase (MCK) cDNA sequence is shown in SEQ ID NO:3, and its deduced amino

acid sequence is shown in SEQ ID NO:4. The binding sites of the gene specific primers for promoter amplification, MCK1 and MCK2, are indicated. The extra nucleotides introduced into MCK1 and MCK2 for generation of restriction sites are shown as a misc_feature in the primer sequences SEQ ID NOS:12 and 13, respectively. A potential polyadenylation signal, AATAAA, is indicated in SEQ ID NO:3.

[0034] The complete fast skeletal muscle isoform of myosin light chain 2 (MLC2f) cDNA sequence is shown in SEQ ID NO:20, and its deduced amino acid sequence is shown in SEQ ID NO:21. The binding sites of the gene-specific primers for promoter amplification, M1 and M2, are indicated. Two potential polyadenylation signals, AATAAA, are shown as a misc_feature in SEQ ID NO:20.

[0035] The complete acidic ribosomal protein P0 (ARP) cDNA sequence is shown in SEQ ID NO:5, and its deduced amino acid sequence is shown in SEQ ID NO:6. The binding sites of the gene specific primers for promoter amplification, ARP1 and ARP2, are indicated. The extra nucleotides introduced into ARP2 for generation of a restriction site are shown as a misc_feature in the primer sequence SEQ ID NO:15. A potential polyadenylation signal, AATAAA, is in-

dicated in SEQ ID NO:5.

[0036] SEQ ID NO:7 shows the complete sequence of the CK promoter region. A putative TATA box is shown, and the 3' nucleotides identical to the 5' CK cDNA sequence are shown as a misc_feature. The binding site of the second gene specific primer, CK2, is shown. The introduced BamHI site is indicated as a misc_feature in the primer sequence SEQ ID NO:11.

[0037] SEQ ID NO:8 shows the complete sequence of the MCK promoter region. A putative TATA box is shown, and the 3' nucleotides identical to the 5' MCK cDNA sequence are shown as a misc_feature in SEQ ID NO:8. The binding site of the second gene specific primer, MCK2, is shown. The introduced BamHI site is indicated as a misc_feature in the primer sequence SEQ ID NO:13.

[0038] SEQ ID NO:22 shows the complete sequence of the MLC2f promoter region. A putative TATA box is shown, and the 3' nucleotides identical to the 5' MLC2f cDNA sequence are shown as a misc_feature. The binding site of the second gene-specific primer, M2, is shown. Potential muscle-specific cis-elements, E-boxes and MEF2 binding sites, are also shown. The proximal 1-kb region of the MLC2f promoter was recently published (Xu et al., 1999).

[0039] SEQ ID NO:9 shows the complete sequence of the ARP promoter region including the first intron. The first intron is shown, and the 3' nucleotides identical to the 5' ARP cDNA sequence are shown as misc_features. No typical TATA box is found. The binding site of the second gene specific primer, ARP2, is shown. The introduced BamHI site is indicated as a misc_feature in the primer sequence SEQ ID NO:15.

[0040] *Specifically Exemplified Polypeptides/DNA.* The present invention contemplates use of DNA that codes for various polypeptides and other types of DNA to prepare the gene constructs of the present invention. DNA that codes for structural proteins, such as fluorescent peptides including GFP, EGFP, BFP, EBFP, YFP, EYFP, CFP, ECFP and enzymes (such as luciferase, ß-galactosidase, chloramphenicol acetyltransferase, etc.), and hormones (such as growth hormone etc.), are useful in the present invention. More particularly, the DNA may code for polypeptides comprising the sequences exemplified in SEQ ID NOS:2, 4, 6 and 21. The present invention also contemplates use of particular DNA sequences, including regulatory sequences, such as promoter sequences shown in SEQ ID NOS: 7, 8, 9 and 22 or portions thereof effective as promoters. Finally,

the present invention also contemplates the use of additional DNA sequences, described generally herein or described in the references cited herein, for various purposes.

[0041] *Chimeric Genes.* The present invention also encompasses chimeric genes comprising a promoter described herein operatively linked to a heterologous gene. Thus, a chimeric gene can comprise a promoter of a zebrafish operatively linked to a zebrafish structural gene other than that normally found linked to the promoter in the genome. Alternatively, the promoter can be operatively linked to a gene that is exogenous to a zebrafish, as exemplified by the GFP and other genes specifically exemplified herein. Furthermore, a chimeric gene can comprise an exogenous promoter linked to any structural gene not normally linked to that promoter in the genome of an organism.

[0042] *Variants of Specifically Exemplified Polypeptide.* DNA that codes for variants of the specifically exemplified polypeptides are also encompassed by the present invention. Possible variants include allelic variants and corresponding polypeptides from other organisms, particularly other organisms of the same species, genus or family. The vari-

ants may have substantially the same characteristics as the natural polypeptides. The variant polypeptide will possess the primary property of concern for the polypeptide. For example, the polypeptide will possess one or more or all of the primary physical (e.g., solubility) and/or biological (e.g., enzymatic activity, physiologic activity or fluorescence excitation or emission spectrum) properties of the reference polypeptide. DNA of the structural genes of the present invention will encode a protein that produces a fluorescent or chemiluminescent light under conditions appropriate to the particular polypeptide in one or more tissues of a fish. Preferred tissues for expression are skin, muscle, eye and bone.

[0043] *Substitutions, Additions and Deletions.* As possible variants of the above specifically exemplified polypeptides, the polypeptide may have additional individual amino acids or amino acid sequences inserted into the polypeptide in the middle thereof and/or at the N-terminal and/or C-terminal ends thereof so long as the polypeptide possesses the desired physical and/or biological characteristics. Likewise, some of the amino acids or amino acid sequences may be deleted from the polypeptide so long as the polypeptide possesses the desired physical and/or

biochemical characteristics. Amino acid substitutions may also be made in the sequences so long as the polypeptide possesses the desired physical and biochemical characteristics. DNA coding for these variants can be used to prepare gene constructs of the present invention.

[0044] *Sequence Identity.* The variants of polypeptides or polynucleotides contemplated herein should possess more than 75% sequence identity (sometimes referred to as homology), preferably more than 85% identity, most preferably more than 95% identity, even more preferably more than 98% identity to the naturally occurring and/or specifically exemplified sequences or fragments thereof described herein. To determine this homology, two sequences are aligned so as to obtain a maximum match using gaps and inserts.

[0045] Two sequences are said to be "identical" if the sequence of residues is the same when aligned for maximum correspondence as described below. The term "complementary" applies to nucleic acid sequences and is used herein to mean that the sequence is complementary to all or a portion of a reference polynucleotide sequence.

[0046] Optimal alignment of sequences for comparison can be conducted by the local homology algorithm of Smith and

Waterman (1981), by the homology alignment method of Needleman and Wunsch (1970), by the search for similarity method of Pearson and Lippman (1988), or the like. Computer implementations of the above algorithms are known as part of the Genetics Computer Group (GCG) Wisconsin Genetics Software Package (GAP, BESTFIT, BLASTA, FASTA and TFASTA), 575 Science Drive, Madison, WI. These programs are preferably run using default values for all parameters.

[0047] "Percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the sequence in the comparison window may comprise additions or deletions (i.e. "gaps") as compared to the reference sequence for optimal alignment of the two sequences being compared. The percentage identity is calculated by determining the number of positions at which the identical residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window and multiplying the result by 100 to yield the percentage of sequence identity. Total identity is then determined as the average identity over all of the windows that cover the complete

query sequence.

[0048] *Fragments of Polypeptide.* Genes which code for fragments of the full length polypeptides such as proteolytic cleavage fragments which contain at least one, and preferably all, of the abovephysical and/or biological properties are also encompassed by the present invention.

[0049] *DNA and RNA.* The invention encompasses DNA that codes for any one of the abovepolypeptides including, but not limited to, those shown in SEQ ID NOS:2, 4, 6 and 21 including fusion polypeptides, variants and fragments thereof. The sequence of certain particularly useful cDNAs which encode polypeptides are shown in SEQ ID NOS:1, 3, 5 and 20. The present invention also includes cDNA as well as genomic DNA containing or comprising the requisite nucleotide sequences as well as corresponding RNA and antisense sequences.

[0050] Cloned DNA within the scope of the invention also includes allelic variants of the specific sequences presented in the attached Sequence Listing. An "allelic variant" is a sequence that is a variant from that of the exemplified nucleotide sequence, but represents the same chromosomal locus in the organism. In addition to those which occur by normal genetic variation in a population and per-

haps fixed in the population by standard breeding methods, allelic variants can be produced by genetic engineering methods. A preferred allelic variant is one that is found in a naturally occurring organism, including a laboratory strain. Allelic variants are either silent or expressed. A silent allele is one that does not affect the phenotype of the organism. An expressed allele results in a detectable change in the phenotype of the trait represented by the locus.

[0051] A nucleic acid sequence "encodes" or "codes for" a polypeptide if it directs the expression of the polypeptide referred to. The nucleic acid can be DNA or RNA. Unless otherwise specified, a nucleic acid sequence that encodes a polypeptide includes the transcribed strand, the hnRNA and the spliced RNA or the DNA representative of the mRNA. An "antisense" nucleic acid is one that is complementary to all or part of a strand representative of mRNA, including untranslated portions thereof.

[0052] *Degenerate Sequences.* In accordance with degeneracy of genetic code, it is possible to substitute at least one base of the base sequence of a gene by another kind of base without causing the amino acid sequence of the polypeptide produced from the gene to be changed. Hence, the

DNA of the present invention may also have any base sequence that has been changed by substitution in accordance with degeneracy of genetic code.

[0053] *DNA Modification.* The DNA is readily modified by substitution, deletion or insertion of nucleotides, thereby resulting in novel DNA sequences encoding the polypeptide or its derivatives. These modified sequences are used to produce mutant polypeptide and to directly express the polypeptide. Methods for saturating a particular DNA sequence with random mutations and also for making specific site-directed mutations are known in the art; see e.g. Sambrook et al. (1989).

[0054] *Hybridizable Variants.* The DNA molecules useful in accordance with the present invention can comprise a nucleotide sequence selected from the group consisting of SEQ ID NOS.:1, 3, 5, 7–20 and 22–24 or can comprise a nucleotide sequence that hybridizes to a DNA molecule comprising the nucleotide sequence of SEQ ID NOS.:1, 3, 5 or 20 under salt and temperature conditions providing stringency at least as high as that equivalent to 5x SSC and 42°C and that codes on expression for a polypeptide that has one or more or all of the abovephysical and/or biological properties. The present invention also includes

polypeptides coded for by these hybridizable variants. The relationship of stringency to hybridization and wash conditions and other considerations of hybridization can be found in Chapters 11 and 12 of Sambrook et al (1989). The present invention also encompasses functional promoters which hybridize to SEQ ID NOS:7, 8, 9 or 22 under the above-described conditions. DNA molecules of the invention will preferably hybridize to reference sequences under more stringent conditions allowing the degree of mismatch represented by the degrees of sequence identity enumerated above. The present invention also encompasses functional primers or linker oligonucleotides set forth in SEQ ID NOS:10-19 and 23-24 or larger primers comprising these sequences, or sequences which hybridize with these sequences under the above-described conditions. The primers usually have a length of 10-50 nucleotides, preferably 15-35 nucleotides, more preferably 18-30 nucleotides.

[0055] *Vectors.* The invention is further directed to a replicable vector containing cDNA that codes for the polypeptide and that is capable of expressing the polypeptide.

[0056] The present invention is also directed to a vector comprising a replicable vector and a DNA sequence corresponding

to the above described gene inserted into said vector. The vector may be an integrating or nonvector depending on its intended use and is conveniently a plasmid.

[0057] *Transformed Cells.* The invention further relates to a transformed cell or microorganism containing cDNA or a vector which codes for the polypeptide or a fragment or variant thereof and that is capable of expressing the polypeptide.

[0058] *Expression Systems Using Vertebrate Cells.* Interest has been great in vertebrate cells, and propagation of vertebrate cells in culture (tissue culture) has become a routine procedure. Examples of vertebrate host cell lines useful in the present invention preferably include cells from any of the fish described herein. Expression vectors for such cells ordinarily include (if necessary) an origin of replication, a promoter located upstream from the gene to be expressed, along with a ribosome-binding site, RNA splice site (if introngenomic DNA is used or if an intron is necessary to optimize expression of a cDNA), a polyadenylation site, and a transcription termination sequence.

EXAMPLES

[0059] The following examples are provided by way of illustration only and not by way of limitation. Those of skill will readily recognize a variety of noncritical parameters which can

be changed or modified to yield essentially similar results.

[0060] *Example I: Isolation of skin-specific, muscle-specific and ubiquitously expressed zebrafish cDNA clones.* cDNA clones were isolated and sequenced as described by Gong et al. (1997). Basically, random cDNA clones were selected from zebrafish embryonic and adult cDNA libraries and each clone was partially sequenced by a single sequencing reaction. The partial sequences were then used to identify the sequenced clones for potential function and tissue specificity. Of the distinct clones identified by this approach, four of them were selected: for skin specificity (clone A39 encoding cytokeratin, CK), for muscle specificity (clone E146 encoding muscle creatine kinase, MCK), for skeletal muscle specificity (clone A113 encoding the fast skeletal muscle isoform of the myosin light chain 2, MLC2f) and for ubiquitous expression (clone A150 encoding acidic ribosomal protein P0, ARP), respectively.

[0061] The four cDNA clones were sequenced, and their complete cDNA sequences with deduced amino acid sequences are shown in SEQ ID NOS:1, 3, 5, and 20 respectively. A39 encodes a type II basic cytokeratin and its closest homolog in mammals is cytokeratin 8 (65-68% amino acid identity). E146 codes for the zebrafish MCK and its amino acid se-

quence shares ~87% identity with mammalian MCKs. A113 encodes the fast skeletal muscle isoform of the myosin light chain 2. The deduced amino acid sequence of this gene is highly homologous to other vertebrate fast skeletal muscle MLC2f proteins (over 80% amino acid identity). The amino acid sequence of zebrafish ARP deduced from the A150 clone is 87–89% identical to those of mammalian ARPs.

[0062] To demonstrate their expression patterns, whole mount in situ hybridization (Thisse et al., 1993) was performed for developing embryos and Northern blot analyses (Gong et al., 1992) were carried out for selected adult tissues and for developing embryos.

[0063] As indicated by whole mount in situ hybridization, cytokeratin mRNA was specifically expressed in the embryonic surface (Figs. 1A–1C) and cross section of in situ hybridized embryos confirmed that the expression was only in skin epithelia (Fig. 1C). Ontogenetically, the cytokeratin mRNA appeared before 4 hours post-fertilization (hpf) and it is likely that the transcription of the cytokeratin gene starts at mid-blastula transition when the zygotic genome is activated. By in situ hybridization, a clear cytokeratin mRNA signal was detected in highly flattened

cells of the superficial layer in blastula and the expression remained in the superficial layer which eventually developed into skin epithelia including the yolk sac. In adult tissues, cytokeratin mRNA was predominantly detected in the skin and also weakly in several other tissues including the eye, gill, intestine and muscle, but not in the liver and ovary (Fig. 2). Therefore, the cytokeratin mRNA is predominantly, if not specifically, expressed in skin cells.

[0064] MCK mRNA was first detected in the first few anterior somites in 10 somite stage embryos (14 hpf) and at later stages the expression is specifically in skeletal muscle (Fig. 1D) and in heart (data not shown). When the stained embryos are cross-sectioned, the MCK mRNA signal was found exclusively in the trunk skeletal muscles (Fig. 1E). In adult tissues, MCK mRNA was detected exclusively in the skeletal muscle (Fig. 2).

[0065] MLC2f mRNA was specifically expressed in fast skeletal muscle in developing zebrafish embryos (Figs.1H-1I). To examine the tissue distribution of MLC2f mRNA, total RNAs were prepared from several adult tissues including heart, brain, eyes, gills, intestine, liver, skeletal muscle, ovary, skin, and testis. MLC2f mRNA was only detected in the skeletal muscle by Northern analysis; while α-actin

mRNA was detected ubiquitously in the same set of RNAs, confirming the validity of the assay (Fig. 2B).

[0066] ARP mRNA was expressed ubiquitously and it is presumably a maternal mRNA since it is present in the ovary as well as in embryos at one cell stage. In in situ hybridization experiments, an intense hybridization signal was detected in most tissues. An example of a hybridized embryo at 28 hpf is shown in Fig. 1F. In adults, ARP mRNA was abundantly expressed in all tissues examined except for the brain where a relatively weak signal was detected (Fig. 2A). These observations confirmed that the ARP mRNA is expressed ubiquitously.

[0067] *Example II:Isolation of zebrafish gene promoters*Four zebrafish gene promoters were isolated by a linker-mediated PCR method as described by Liao et al., (1997) and as exemplified by the diagrams in Fig. 3. The whole procedure includes the following steps: 1) designing of gene specific primers; 2) isolation of zebrafish genomic DNA; 3) digestion of genomic DNA by a restriction enzyme; 4) ligation of a short linker DNA to the digested genomic DNA; 5) PCR amplification of the promoter region; and 6) DNA sequencing to confirm the cloned DNA fragment. The following is the detailed description of these steps.

[0068] 1. Designing of gene specific primers. Gene specific PCR primers were designed based on the 5' end of the four cDNA sequences and the regions used for designing the primers are shown in SEQ ID NOS: 1, 3, 5 and 20.

[0069] The two cytokeratin gene specific primers are:CK1 (SEQ ID NO:10)CK2 (SEQ ID NO:11), where the first six nucleotides are for creation of an EcoRI site to facilitate cloning.

[0070] The two muscle creatine kinase gene specific primers are:MCK1 (SEQ ID NO:12), where the first five nucleotides are for creation of an EcoRI site to facilitate cloning.

[0071] MCK2 (SEQ ID NO:13), where the first three nucleotides are for creation of an EcoRI site to facilitate cloning.

[0072] The two fast skeletal muscle isoform of myosin light chain 2 gene specific primers are:M1 (SEQ ID NO:23) M2 (SEQ ID NO:24)The two acidic ribosomal protein P0 gene specific primers are:ARP1 (SEQ ID NO:14)ARP2 (SEQ ID NO:15), where the first six nucleotides are for creation of an EcoRI site to facilitate cloning.

[0073] 2. Isolation of zebrafish genomic DNA. Genomic DNA was isolated from a single individual fish by a standard method (Sambrook et al., 1989). Generally, an adult fish was quickly frozen in liquid nitrogen and ground into powder. The ground tissue was then transferred to an ex-

traction buffer (10 mM Tris, pH 8, 0.1 M EDTA, 20 µg/ml RNase A and 0.5% SDS) and incubated at 37°C for 1 hour. Proteinase K was added to a final concentration of 100 µg/ml and gently mixed until the mixture appeared viscous, followed by incubation at 50°C for 3 hours with periodical swirling. The genomic DNA was gently extracted three times by phenol equilibrated with Tris-HCl (pH 8), precipitated by adding 0.1 volume of 3 M NaOAc and 2.5 volumes of ethanol, and collected by swirling on a glass rod, then rinsed in 70% ethanol.

[0074] 3. Digestion of genomic DNA by a restriction enzyme. Genomic DNA was digested with the selected restriction enzymes. Generally, 500 units of restriction enzyme were used to digest 50 µg of genomic DNA overnight at the optimal enzyme reaction temperature (usually at 37°C).

[0075] 4. Ligation of a short linker DNA to the digested genomic DNA. The linker DNA was assembled by annealing equal moles of the two linker oligonucleotides, Oligo1 (SEQ ID NO:16) and Oligo 2 (SEQ ID NO:17). Oligo 2 was phosphorylated by T4 polynucleotide kinase prior to annealing. Restriction enzyme digested genomic DNA was filled-in or trimmed with T4 DNA polymerase, if necessary, and ligated with the linker DNA. Ligation was performed with 1

µg of digested genomic DNA and 0.5 µg of linker DNA in a 20 µl reaction containing 10 units of T4 DNA ligase at 4°C overnight.

[0076] 5. PCR amplification of promoter region. PCR was performed with Advantage Tth Polymerase Mix (Clontech). The first round of PCR was performed using a linker specific primer L1 (SEQ ID NO:18) and a gene specific primer G1 (CK1, MCK1, M1 or ARP1). Each reaction (50 µl) contains 5 µl of 10x Tth PCR reaction buffer (1X= 15 mM KOAc, 40 mM Tris, pH 9.3), 2.2 µl of 25 mM Mg(OAc)2, 5 µl of 2 mM dNTP, 1 µl of L1 (0.2 µg/µl), 1 µl of G1 (0.2 µg/µl) , 33.8 µl of H2O, and 1 µl (50 ng) of linker ligated genomic DNA and 1 µl of 50x Tth polymerase mix (Clontech). The cycling conditions were as follows: 94°C/1 min, 35 cycles of 94°C/30 sec and 68°C/6 min, and finally 68°C/8 min. After the primary round of PCR was completed, the products were diluted 100 fold. One µl of diluted PCR product was used as template for the second round of PCR (nested PCR) with a second linker specific primer L2 (SEQ ID NO:19) and a second gene specific primer G2 (CK2, MCK2, M2 or ARP2), as described for the primary PCR but with the following modification: 94°C/1 min, 25 cycles of 94°C/30 sec and 68°C/6 min, and finally

68°C/8 min. Both the primary and secondary PCR prod-
ucts were analyzed on a 1% agarose gel.

[0077] 6. DNA sequencing to confirm the cloned DNA fragment.
PCR products were purified from the agarose gel following
electrophoresis and cloned into a TA vector,
pT7Blue™(Novogen). DNA sequencing was performed by
dideoxynucleotide chain termination method using a T7
Sequencing Kit purchased from Pharmacia. Complete se-
quences of these promoter regions were obtained by au-
tomatic sequencing using a dRhodamine Terminator Cycle
Sequencing Ready Reaction Kit (Perkin-Elmer) and an ABI
377 automatic sequencing machine.

[0078] The isolated cytokeratin DNA fragment comprising the
gene promoter is 2.2 kb. In the 3' proximal region imme-
diately upstream of a portion identical to the 3' part of the
CK cDNA sequence, there is a putative TATA box perfectly
matching to a consensus TATA box sequence. The 164 bp
of the 3' region is identical to the 5' UTR (untranslated re-
gion) of the cytokeratin cDNA. Thus, the isolated fragment
was indeed derived from the same gene as the cytokeratin
cDNA clone (SEQ ID NO:7). Similarly, a 1.5 kb 5' flanking
region was isolated from the muscle creatine kinase gene,
a putative TATA box was also found in its 3' proximal re-

gion and the 3' region is identical to the 5' portion of the MCK cDNA clone (SEQ ID NO:8). For MLC2f, a 2 kb region was isolated from the fast skeletal muscle isoform of myosin light chain 2 gene and sequenced completely. The promoter sequence for MLC2f is shown in SEQ ID NO:22. The sequence immediately upstream of the gene specific primer M2 is identical to the 5' UTR of the MLC2f cDNA clone; thus, the amplified DNA fragments are indeed derived from the MLC2f gene. A perfect TATA box was found 30 nucleotides upstream of the transcription start site, which was defined by a primer extension experiment based on Sambrook et al. (1989). In the 2–kb region comprising the promoter, six E–boxes (CANNTG) and six potential MEF2 binding sites [C/T)TA(T/A)4TA(A/G)] were found and are indicated in SEQ ID NO:22. Both of these cis–element classes are important for muscle specific gene transcription (Schwarz et al., 1993; Olson et al., 1995). A 2.2 kb fragment was amplified for the ARP gene. By alignment of its sequence with the ARP cDNA clone, a 1.3 kb intron was found in the 5' UTR (SEQ ID NO:9). As a result, the isolated ARP promoter is within a DNA fragment about 0.8 kb long.

[0079] *Example III: Generation of green fluorescent transgenic fish* The

isolated zebrafish gene promoters were inserted into the plasmid pEGFP-1 (Clonetech), which contains an EGFP structural gene whose codons have been optimized according to preferable human codons. Three promoter fragments were inserted into pEGFP-1 at the EcoRI and BamHI site and the resulting recombinant plasmids were named pCK-EGFP (Fig. 4), pMCK-EGFP (Fig. 5), and pARP-EGFP, respectively (Fig. 6). The promoter fragment for the MLC2f gene was inserted into the Hind III and Bam HI sites of the plasmid pEGFP-1 and the resulting chimeric DNA construct, pMLC2f-EGFP, is diagramed in Fig. 7.

[0080] Linearized plasmid DNAs at a concentrations of 500 µg/ml (for pCK-EGFP and pMCK-EGFP) and 100 µg/ml (for pMLC2f-EGFP) in 0.1 M Tris-HCl (pH 7.6)/0.25% phenol red were injected into the cytoplasm of 1- or 2-cell stage embryos. Because of a high mortality rate, pARP-EGFP was injected at a lower concentration (50 µg/ml). Each embryo received 300-500 pl of DNA. The injected embryos were reared in autoclaved Holtfreter"s solution (0.35% NaCl, 0.01% KCl and 0.01% CaCl2) supplemented with 1 µg/ml of methylene blue. Expression of GFP was observed and photographed under a ZEISS Axiovert 25 fluorescence microscope.

[0081]   When zebrafish embryos received pCK–EGFP, GFP expression started about 4 hours after injection, which corresponds to the stage of ~30% epiboly. About 55% of the injected embryos expressed GFP at this stage. The early expression was always in the superficial layer of cells, mimicking endogenous expression of the CK gene as observed by in situ hybridization. At later stages, in all GFP-expressing fish, GFP was found predominantly in skin epithelia. A typical pCK–EGFP transgenic zebrafish fry at 4 days old is shown in Fig. 8.

[0082]   Under the MCK promoter, no GFP expression was observed in early embryos before muscle cells become differentiated. By 24 hpf, about 12% of surviving embryos expressed GFP strongly in muscle cells and these GFP-positive embryos remain GFP-positive after hatching. The GFP expression was always found in many bundles of muscle fibers, mainly in the mid-trunk region and no expression was ever found in other types of cells. A typical pMCK–EGFP transgenic zebrafish fry (3 days old) is shown in Fig. 9.

[0083]   Expression of pARP–EGFP was first observed 4 hours after injection at the 30% epiboly stage. The timing of expression is similar to that of pCK–EGFP–injected embryos.

However, unlike the pCK-EGFP transgenic embryos, the GFP expression under the ARP promoter occurred not only in the superficial layer of cells but also in deep layers of cells. In some batches of injected embryos, almost 100% of the injected embryos expressed initially. At later stages when some embryonic cells become overtly differentiated, it was found that the GFP expression occurred essentially in all different types of cells such as skin epithelia, muscle cells, lens, neural tissues, notochord, circulating blood cells and yolk cells (Fig. 10).

[0084] Under the MLC2f promoter, nearly 60% of the embryos expressed GFP. The earliest GFP expression started in trunk skeletal muscles about 19 hours after injection, which corresponds to the stage of 20-somite. Later, the GFP expression also occurred in head skeletal muscles including eye muscles, jaw muscles, gill muscles etc.

[0085] Transgenic founder zebrafish containing pMLC2f-EGFP emit a strong green fluorescent light under a blue or ultraviolet light (Fig. 11A). When the transgenic founders were crossed with wild-type fish, transgenic offspring were obtained that also displayed strong green fluorescence (Fig. 11B). The level of GFP expression is so high in the transgenic founders and offspring that green fluores-

cence can be observed when the fish are exposed to sun-light.

[0086] To identify the DNA elements conferring the strong promoter activity in skeletal muscles, deletion analysis of the 2-kb DNA fragment comprising the promoter was performed. Several deletion constructs, which contain 5" deletions of the MLC2f promoter upstream of the EGFP gene, were injected into the zebrafish embryos and the transient expression of GFP in early embryos (19–72 hpf) was compared. To facilitate the quantitative analysis of GFP expression, we define the level of expression as follows (Figs. 12A–12C):Strong expression: GFP expression was detected in essentially 100% muscle fibers in the trunk.

[0087] Moderate expression: GFP expression was detected in several bundles of muscle fibers, usually in the mid-trunk region.

[0088] Weak expression: GFP expression occurred in dispersed muscle fibers and the number of GFP positive fibers is usually less than 20 per embryo.

[0089] As shown in Fig. 13, deletion up to 283 bp maintained the GFP expression in skeletal muscles in 100% of the expressing embryos; however, the level of GFP expression

from these deletion constructs varies greatly. Strong expression drops from 23% to 0% from the 2-kb (-2011 bp) promoter to the 283-bp promoter. Thus, only two constructs (2011 bp and 1338 bp) are capable of maintaining the high level of expression and the highest expression was obtained only with the 2-kb promoter, indicating the importance of the promoter region of 1338 bp to 2011 bp for conferring the highest promoter activity.

[0090] The expression of GFP using pMLC2f-EGFP is much higher than that obtained using the pMCK-EGFP that contains a 1.5 kb of zebrafish MCK promoter (Singapore Patent Application 9900811-2). By the same assay in transient transgenic zebrafish embryos, only about 12% of the embryos injected with pMCK-EGFP expressed GFP. Among the expressing embryos, no strong expression was observed, and 70% and 30% showed moderate and weak expression, respectively. In comparison, about 60% of the embryos injected with pMLC2f-EGFP expressed GFP and 23%, 37% and 40% showed strong, moderate and weak expression, respectively.

[0091] *Example IV:Potential applications of fluorescent transgenic fish*The fluorescent transgenic fish have use as ornamental fish in the market. Stably transgenic lines can be developed by

breeding a GFP transgenic individual with a wild type fish or another transgenic fish. By isolation of more zebrafish gene promoters, such as eye-specific, bone-specific, tail-specific etc., and/or by classical breeding of these transgenic zebrafish, more varieties of fluorescent transgenic zebrafish can be produced. Previously, we have reported isolation of over 200 distinct zebrafish cDNA clones homologous to known genes (Gong et al., 1997). These isolated clones code for proteins in a variety of tissues and some of them are inducible by heat-shock, heavy metals, or hormones such as estrogens. By using the method of PCR amplification using gene-specific primers designed from the nucleotide sequences of these cDNAs, and the linker-specific primers described herein, the promoters of the genes represented by the cDNAs of Gong et al. can be used in the present invention. Thus, hormone-inducible promoters, heavy-metal inducible promoters and the like from zebrafish can be isolated and used to make fluorescent zebrafish (or other fish species) that express a GFP or variant thereof, in response to the relevant compound.

[0092] Multiple color fluorescent fish may be generated by the same technique as blue fluorescent protein (BFP) gene, yellow fluorescent protein (YFP) gene and cyan fluorescent

protein (CFP) gene are available from Clonetech. For example, a transgenic fish with GFP under an eye-specific promoter, BFP under a skin-specific promoter, and YFP under a muscle-specific promoter will show the following multiple fluorescent colors: green eyes, blue skin and yellow muscle. By recombining different tissue specific promoters and fluorescent protein genes, more varieties of transgenic fish of different fluorescent color patterns will be created. By expression of two or more different fluorescent proteins in the same tissue, an intermediate color may be created. For example, expression of both GFP and BFP under a skin-specific promoter, a dark-green skin color may be created.

[0093] By using a heavy metal- (such as cadmium, cobalt, chromium) inducible or hormone- (such as estrogen, androgen or other steroid hormone) inducible promoter, a biosensor system may be developed for monitoring environmental pollution and for evaluating water quality for human consumption and aquacultural uses. In such a biosensor system, the transgenic fish will glow with a green fluorescence (or other color depending on the fluorescence protein gene used) when pollutants such as heavy metals and estrogens (or their derivatives) reach a

threshold concentration in an aquatic environment. Such a biosensor system has advantages over classical analytical methods because it is rapid, visualizable, and capable of identifying specific compounds directly in complex mixture found in an aquatic environment, and is portable or less instrument dependent. Moreover, the biosensor system also provides direct information on biotoxicity and it is biodegradable and regenerative.

[0094] Environmental monitoring of several substances can be accomplished by either creating one transgenic fish having genes encoding different colored fluorescent proteins driven by promoters responsive to each substance. Then the particular colors exhibited the fish in an environment can be observed. Alternatively, a number of fish can be transformed with individual vectors, then the fish can be combined into a population for monitoring an environment and the colors expressed by each fish observed.

[0095] In addition, the fluorescent transgenic fish should also be valuable in the market for scientific research tools because they can be used for embryonic studies such as tracing cell lineage and cell migration. Cells from transgenic fish expressing GFP can also be used as cellular and genetic markers in cell transplantation and nuclear trans-

plantation experiments.

[0096] The chimeric gene constructs demonstrated successfully in zebrafish in the present invention should also be applicable to other fish species such as medaka, goldfish, carp including koi, loach, tilapia, glassfish, catfish, angel fish, discus, eel, tetra, goby, gourami, guppy, Xiphophorus (swordtail), hatchet fish, Molly fish, pangasius, etc. The promoters described herein can be used directly in these fish species. Alternatively, the homologous gene promoters from other fish species can be isolated by the method described in this invention. For example, the isolated and characterized zebrafish cDNA clones and promoters described in this invention can be used as molecular probes to screen for homologous promoters in other fish species by molecular hybridization or by PCR. Alternatively, one can first isolate the zebrafish cDNA and promoters based on the sequences presented in SEQ ID NOS:1, 3, 5, 7, 8, 9, 20 and 22 or using data from other sequences of cDNAs disclosed by Gong et al. 1997, by PCR and then use the zebrafish gene fragments to obtain homologous genes from other fish species by the methods mentioned above.

[0097] In addition, a strong muscle-specific promoter such as MLC2f is valuable to direct a gene to be expressed in

muscle tissues for generation of other beneficial trans-
genic fish. For example, transgenic expression of a
growth hormone gene under the muscle-specific pro-
moter may stimulate somatic growth of transgenic fish.
Such DNA can be introduced either by microinjection,
electroporation, or sperm carrier to generate germ-line
transgenic fish, or by direct injection of naked DNA into
skeletal muscles (Xu et al., 1999) or into other tissues or
cavities, or by a biolistic method (gene bombardment or
gene gun) (Gomez-Chiarri et al., 1996).